Title

Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest.

Journal

Nature neuroscience, 17(10)

ISSN

1097-6256

Authors

Zhu, Lusha
Jenkins, Adrianna C
Set, Eric
et al.

Publication Date

2014-10-01

DOI

10.1038/nn.3798

Peer reviewed

HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2015 April 01.

Published in final edited form as:

Nat Neurosci. 2014 October ; 17(10): 1319–1321. doi:10.1038/nn.3798.

Damage To Dorsolateral Prefrontal Cortex Affects Tradeoffs Between Honesty And Self-Interest

Lusha Zhu[1], **Adrianna C. Jenkins**[2], **Eric Set**[2,3], **Donatella Scabini**[4,5], **Robert T. Knight**[4,5], **Pearl H. Chiu**[1,6,7], **Brooks King-Casas**[1,6,7,8], and **Ming Hsu**[2,5,*]

[1]Virginia Tech Carilion Research Institute, Roanoke, VA 24016

[2]Haas School of Business, University of California, Berkeley, Berkeley CA 94720

[3]Department of Economics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[4]Department of Psychology, University of California, Berkeley, Berkeley CA 94720

[5]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley CA 94720

[6]Department of Psychology, Virginia Tech, Blacksburg, VA 24061

[7]Department of Psychiatry, Virginia Tech Carilion School of Medicine, Roanoke, VA 24016

[8]Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences, Blacksburg, VA 24061

Abstract

Substantial correlational evidence exists suggesting a critical role for prefrontal regions in honest and dishonest behavior, but causal evidence specifying the nature of this involvement remains absent. Here we show using the lesion method that damage to the human dorsolateral prefrontal cortex (DLPFC) decreased the effect of honesty concerns on behavior in economic games that pit honesty motives against self-interest, but did not affect decisions where honesty concerns were absent. These results point to a causal role for DLPFC in enabling honest behavior.

A wealth of field and laboratory studies have shown that humans are often willing to sacrifice their own economic payoffs in the interest of being honest, even in the absence of punishment or reputational factors[1, 2]. At the neural level, there is substantial evidence from both neuroimaging[3–6] and developmental[7, 8] literatures that the prefrontal cortices, in particular dorsolateral prefrontal (DLPFC) and orbitofrontal (OFC) cortices, play a critical role in decisions involving honesty. Due to the inherently correlational nature of such data, however, the specific role of these regions in honesty and dishonesty remains unclear. Here we sought to characterize the causal contribution of these regions by comparing the behavior of patients with focal lesions to either the DLPFC or OFC to that of healthy comparison

Corresponding author: mhsu@haas.berkeley.edu.

AUTHOR CONTRIBUTIONS

L.Z., E.S., P.H.C., B.K.C., and M.H. designed the experiments, E.S. and D.S. carried out the experiments, L.Z., A.C.J., E.S., R.T.K., and M.H. carried out statistical analyses, and all authors wrote the paper.

participants in a battery of signaling games extensively studied in behavioral economics and evolutionary biology[9, 10] (Fig. 1, Supplementary Fig. 1-2, Supplementary Table 1, Online Methods). These games capture a core dilemma involved in honest behavior where interests of the signaler conflicts with those of the signal receiver, such as that of a seller (signaler) choosing to either truthfully disclose or misrepresent information about a product's quality, which has direct monetary consequences for the buyer (signal receiver).

First, in the Message condition, the participant in the role of the signaler can send one of two messages to an anonymous counterpart in the role of the signal recipient, on the basis of which the recipient chooses one of two monetary allocation associated with the messages (Fig. 2a, Online Methods)[2, 10]. Importantly, both players were instructed that only the signaler would be informed about the monetary consequences associated with each option, and that recipients will never know if a message they received was true (Online Methods). This highlights the fact that the signal recipient is entirely reliant upon the signaler for potential information about the options, and prevents them from using payoff information to make inferences about signaler behavior[2, 10].

In addition, to account for possible baseline differences in altruistic tendencies, we included a Choice condition that contained matching monetary consequences to those in the Message condition (Online Methods). The only difference between the conditions is that, in the Choice condition, participants directly chose between Option A and Option B. An individual who is completely insensitive to honesty concerns will behave identically in the two conditions, whereas those sensitive to honesty concerns are predicted to behave more generously in the Message condition. All choices were conducted using hypothetical payoffs and no feedback, with order of Message and Choice blocks counterbalanced across participants (Online Methods, Supplementary Table 2).

We first investigated how introduction of honesty concerns affected choice behavior in healthy participants by comparing altruistic giving in the Message and Choice conditions, defined as the amount received by the recipient following implementation of the participant's decision, which for simplicity we refer to as "amount given" (Online Methods). Using paired comparisons on decisions with identical monetary consequences, we found that consistent with previous studies in healthy participants[1, 2, 10], inclusion of honesty concerns in the Message condition substantially increased altruistic giving compared to the Choice condition (Wilcoxon signed-rank test, $p < .001$, two-tailed; Fig. 2b).

To test the extent to which prefrontal regions are causally involved in trade-offs between honesty concerns and economic self-interest, we next compared amount given between the Message and Choice conditions in patients with lesions to either DLPFC or OFC versus healthy participants. We found significant main effects of both condition (Wilcoxon signed-rank test, $p < .001$, two-tailed), such that participants on average gave more in the Message condition ($7.90 ± .20) than in the Choice condition ($4.14 ± .24), and cohort (Kruskal-Wallis test, $p < .001$, two-tailed), such that DLPFC patients ($7.17 ± .33) on average gave less than healthy participants ($8.91 ± .19) and OFCs ($8.30 ± .35). Critically, we observed a significant interaction between condition and cohort (Kruskal-Wallis test on paired difference in amount given across 3 cohorts, $p < .001$, two-tailed), such that damage to

DLPFC was associated with significantly lower giving amounts than other cohorts in the Message condition but not in the Choice condition, suggesting a reduction in the sensitivity to honesty concerns without changes in baseline altruistic tendencies on the part of DLPFC patients (Fig. 2b; Supplementary Fig. 3-5). All results are robust to using parametric statistical tests. For additional details on the relationship between behavior and demographic variables and lesion laterality, see Supplementary Figure 3 and Supplementary Table 3.

To assess the possibility that deficits in cognitive processes unrelated to honesty may have produced the observed behavioral differences, we first separated decisions in the Message condition where honesty and self-interest were in conflict from decisions where the two were aligned (Online Methods). If behavioral patterns observed in DLPFC cohort reflected general impairments such as misunderstanding of payoffs or different beliefs about the behavior of the signal recipients, we would expect DLPFC patients to be affected on both types of decisions. In contrast, we found that DLPFC patients were selectively affected in conflict trials (Fig. 2c top) and were indistinguishable from healthy comparison or OFC cohorts in no conflict trials (Fig. 2c bottom; Supplementary Fig. 3b). In addition, we did not find support for the hypothesis that DLPFC patients exhibited more random choice behavior in the Message condition, therefore exerting downward bias on the effect of honesty (Supplementary Fig. 6). For additional behavioral results validating task design, see Supplementary Fig. 7-8.

The above results are thus consistent with previous suggestions that DLPFC influences value computations by diminishing subjective value associated with the pursuit of immediate self-interest[11, 12]. To formally test this mechanistic hypothesis, we used a computational approach to characterize how parametric variation in costs and benefits associated with honesty influenced choice behavior in our different cohorts. Specifically, we assumed the subjective value of an option is influenced not only by monetary consequences to self and other but also the means (honest or dishonest) by which these outcomes are obtained (Online Methods, Supplementary Table 4)[10]. We found that the weight placed on participants' own payoff decreased in the Message condition (α_M) for the OFC and healthy comparison cohorts by approximately 50% relative to the Choice condition (α_C; Fig. 3a). Strikingly, DLPFC patients' choices did not exhibit a significant discrepancy in the weight across two conditions (Fig. 3b), and were significantly different from those of both healthy comparison and OFC cohorts (Fig. 3b).

Together, our findings suggest a necessary role for DLPFC in promoting honesty concerns over self-interested motives, and argue against the widely proposed view that the involvement of prefrontal regions in honesty reflects the need to engage regulatory processes to override truthful responses and implement self-interest[3, 13]. Under the latter hypothesis, damage to prefrontal regions should have been associated with an increased sensitivity to honesty concerns, resulting in greater altruistic tendencies when honesty came into conflict with self-interest. Instead, the current results are consistent with the idea that control is necessary to curb self-interest motives in order to communicate the truth, and further suggest that previous neuroimaging findings of DLPFC engagement during dishonest behavior reflect active, but ultimately unsuccessful, engagement of control processes,

consistent with observations that individuals with control deficits often engage DLPFC more[14, 15].

In contrast to the DLPFC, we did not observe an effect of OFC damage on behavior, which might reflect a number of features of our task, including the reduction of anticipated guilt and lack of strong affective components (Supplementary Fig. 9)[16, 17]. At the same time, we cannot completely rule out possible contributions from non-PFC based processes to honesty due to the presence of damage to white matter and in some cases extending into adjacent regions in our lesion sample (Fig. 1; Supplementary Fig. 1-2). Future studies combining larger lesion cohorts with functional connectivity measures will be needed to address these questions[18]. More broadly, by connecting tools and ideas from behavioral economics and theoretical biology with those of cognitive neuroscience, our study raises exciting questions regarding the degree to which the neurocomputational substrates of honesty are shared with other types of norm-guided and moral behavior[19, 20], as well as regarding the neural mechanisms necessary for arbitrating between such norms in cases of conflict.

Online Methods

Subjects

Patients with focal brain lesions to the dorsolateral prefrontal cortex (n = 7) and orbitofrontal cortex (n = 7) were included in the experiment. Healthy comparison participants (n = 27) were recruited from San Francisco Bay Area, CA. All subjects provided informed consent approved by the University of California, Berkeley, CA. One DLPFC lesion patient answered incorrectly on more than 50% of post instruction questionnaires, and was excluded from the study. In comparison no other subjects failed to answer fewer than 90% of the questions correctly. All statistical results reported in the study are robust to inclusion of this participant.

Table S1

Demographic information and neuropsychological background.

	N	Age	Gender (F)	Years of education	Estimated WAIS[1]	Etiology	Hemisphere
DLPFC	6	57 (8.37)	4	16.17 (2.86)	99 (8.05)	stroke (6)	left (5) right (1)
OFC[2]	7	46.71 (16.86)	3	15.14 (2.85)	109.83 (9.26)	traumatic brain injury[3] (6) tumor resection (1)	bilateral (6) left (1)
Healthy Comparison	27	48.31 (14.40)	12	15.81 (1.10)	105.5 (13.52)	NA	NA

Parentheses contain standard deviations. WAIS: Wechsler Adult Intelligence Scale.

[1] WAIS scores were estimated from Shipley Institute of Living Scale.

[2] The presented WAIS is an average over 6 OFC patients as one patient did not complete the IQ test.

Lesion Reconstruction

Software reconstructions were performed using MRIcron[21]. For both patient groups, testing took place at least 6 months after the date of the stroke/accident. A neurologist (R.T.K.)

inspected patient MRIs to ensure that no white matter hyperintensities outside the lesioned area were observed in either patient group. All TBI patients had low impact force injuries with no clinical or MRI evidence of axonal shear.

Signaling Games

We used a battery of signaling games extensively studied in behavioral economics and evolutionary biology[9, 10]. These games capture a core dilemma involved in honest behavior where interests of the signaler conflicts with those of the signal receiver, such as that of a seller (signaler) choosing to either truthfully disclose or misrepresent information about a product's quality, which has direct monetary consequences for the buyer (signal receiver).

These games have three important advantages as an assay of decisions involving tradeoffs between honesty and self-interest. First, to isolate the effects of honesty, we included a set of Message and Choice conditions. Because the latter condition does not include honesty concerns, we remove the tension between honesty and other social preferences and are able to control for participants' concerns for equity and efficiency. As a result, systematic deviations in behavior between the two sets of games can be interpreted as being affected by honesty concerns. Specifically, an individual who is completely insensitive to honesty concerns will behave identically in the two conditions, whereas those sensitive to honesty concerns are predicted to behave more generously in the Message condition. In previous experiments using these games, introduction of honesty concerns in the Message condition has been found to increase cooperation rates and altruistic giving by approximately 50%[1, 2, 10, 22].

Second, the clearly delineated cost-benefit relationship associated with self-interest and honesty facilitates a computational account of honesty, which allows us to better connect the potential behavioral differences to their computational substrates. Finally, and importantly in the context of lesion studies, by explicitly presenting honest and self-interested actions to subjects, the Message condition allows us to hold constant the available action set across cohorts and verify understanding. This included both comprehension tests and control trials with no conflict between honesty and self-interest.

Message and Choice conditions

In the Message condition, the participant in the role of signaler was presented with two options, A and B, which yielded different monetary outcomes. For example, in Fig. S1a, Option A corresponded to $6 to the signaler and $5 to an anonymous random signal recipient, i.e., ($6, $5), and Option B corresponded to ($5, $10). Only the signaler knew the payoffs associated with the options, and had to send either an honest or dishonest message to an anonymous recipient. The recipient did not know the associated payoffs but had to choose one of the two options. That is, the signaler could either choose to convey the truth, "Option B will earn you more money than Option A", or a falsehood, "Option A will earn you more money than Option B". Importantly, all signalers were informed that recipients would never know the payment information associated with each option and therefore whether senders' messages were true or not.

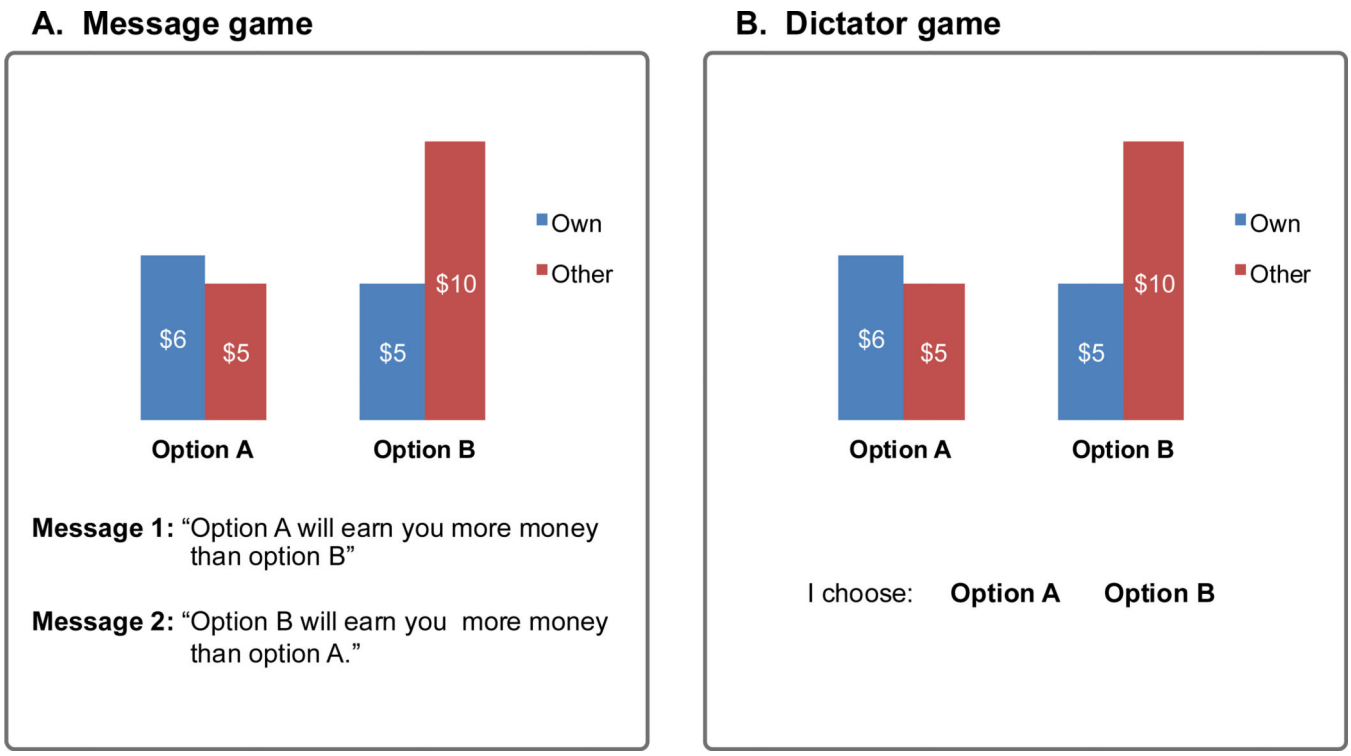


Fig. S1: Task interfaces for **(A)** Message, **(B)** Choice conditions.

The monetary outcomes varied across trials. In particular, in some trials we pit self-interest against honesty. That is, honest choices were associated with allocations that yielded less payment to the participant and more to the recipient (e.g., $5 for self, $15 for other in option A; versus $6 for self, $5 for other in option B). We refer these trials as "conflict trials". In "no conflict trials", honest choices were associated with allocations that yielded more payment to both participant and recipient (e.g., $8 for self, $10 for other in option A; versus $10 for self, $12 for other in option B). Full list of trial options is presented in Supplementary Table 2.

As a control condition, we also included the Choice condition associated with the same set of payoff allocations. In particular, participants were asked to directly choose either Option A or Option B. Following the procedure of previous experiments using the Message and Choice condition[2], participants were informed that in the Choice condition (i) their decisions would be implemented 80% of the time, while the other 20% of the time the alternative option would be implemented; and (ii) receivers would not know the monetary payoff associated with each option and would just receive money passively.

Procedure

Following task instructions and comprehension quiz (Appendix A), participants were administered two blocks of Message and Choice condition trials, each containing 12 trials (Appendix B). All choices were conducted using hypothetical payoffs and no feedback, with order of Message and Choice blocks counterbalanced across participants within each cohort. Within each block, questions were presented in a random order.

Behavioral Analysis

In both conditions, the behavioral measure of altruistic giving was defined as the amount that would be received by the recipient if the participant's decision was implemented, which for simplify we refer to as "amount given". Using payoffs given in Fig. S1 as an example, the amount given in the Message condition by a participant choosing the truthful (false) Message 2 (1) would be defined as $10 ($5). Similarly, in the Choice condition, the amount given by a participant choosing Option A (B) would be defined as $5 ($10).

Computational Modeling

To characterize the relative contributions of economic self-interest, distributional preference, and honesty consideration to allocation decisions, we adapted an economic model that previously applied to study social preferences[23] to our tasks.

First, denote M_S and M_O as monetary payoffs for self and other respectively. The indicator function I is equal to 1 when the monetary payoff is achieved through dishonesty and 0 otherwise. That is, I indicates whether honesty concerns are overridden. We propose that the decision-maker's utility is modulated by honesty in addition to monetary allocations to self and other:

$$U(M_s, M_o, I) = [(\alpha - I \cdot \delta) M_s{}^{\rho} + (1 - \alpha + I \cdot \delta) M_o{}^{\rho}]^{\frac{1}{\rho}}$$

Here α and ρ are parameters capturing distributional preferences that solely depend upon the monetary allocation between self and other, whereas δ quantifies the biasing effects of honesty concerns. The functional form follows the well-established Constant Elasticity of Substitution utility function[24].

Specifically, the parameter α quantifies the relative weight between monetary payoffs for self and other. A large α indicates a larger weight on own economic gain. The parameter ρ reflects the elasticity of substitution between M_S and M_O. For example, if ρ approaches 1, the utility function will reduce to a linear function representing the preference of welfare maximizing. If ρ approaches negative infinity, the utility function will reduce to $U(M_S, M_O, I) = min(M_S, M_O)$, which corresponds to the preference of maximal inequity aversion.

In the context of our game, we refer to α as the weight placed on own payoff in the Choice condition, as there is no tradeoff between self-interest and honesty. That is, $\alpha_C = \alpha$. In contrast, the weight placed on own payoff in the Message condition is defined by $\alpha_M = \alpha - \delta$. Critically, the parameter δ can be interpreted as the degree to which honesty reduces self-interested motives. If $\delta > 0$, the signaler suffers from a disutility of deception and is more likely to sacrifice self-interest in favor of honesty concerns. In contrast, if $\delta < 0$, the signaler receives an additional utility from dishonesty, and thus is more likely to choose dishonest options. Finally, if $\delta = 0$, the signaler is indifferent between honest or dishonest actions and will behave as if the tradeoff between honesty and dishonesty does not exist. The combination of these parameters thus nests a wide range of social preferences proposed by existing theory and allows for rich interactions among economic self-interest, distributional preferences and honesty considerations.

To calibrate the model given the binary choice behavior of each cohort in the game, we adopted the standard logit assumption, aggregated observations conditional on lesion cohorts and experimental conditions and conducted maximal likelihood estimation, specifically maximizing the log likelihood function $\Sigma_i \Sigma_t \log (P_{i,t}(y_{it}; \alpha_M, \alpha_C, \rho))$. The standard errors of estimated parameters were obtained through the bootstrap procedure with 200 iterations for each cohort.

Experiment Instructions

Preamble (delivered verbally)

Imagine that there are anonymous participants in some other rooms that you would never meet and whose identity you would not know, and vice versa. You can assume that none of you are able to communicate with each other. For convenience, we may sometimes refer to these hypothetical participants as “your partner”, but no prior relationship is implied.

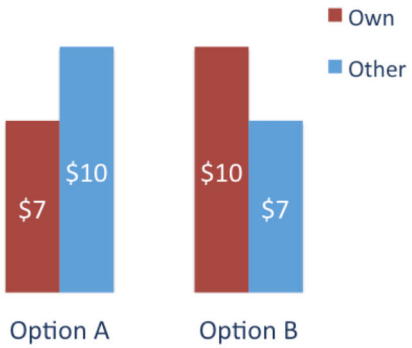
Message Condition

This part of the experiment describes a situation involving you and one other person.

In each decision you will be asked to select between two messages, where the message that you choose will be sent to the other person. After reading your message, the other person will have to make a choice, which will determine the payoffs for both of you.

There are 12 questions altogether. Consider each question separately. That is, treat each decision as if it alone will determine the final payoffs. Although the questions are hypothetical, please do your best to respond as if the decisions are for real money.

Example: In this example, two possible payments are available to you and the other person. The two payment options are:



The other person will have to choose either option A or option B. The only information the other person will have is information sent by you in a message. That is he or she will not know the monetary payments associated with each choice.

We now ask you to choose one of the two possible messages, which you will send to the other person:

Message 1: " Option A will earn you more money than option B."
Message 2: " Option B will earn you more money than option A."

We will show the other person your message, and ask him or her to choose either option A or option B.

To repeat, the other person's choice will determine the payments in the experiment. However, the other person will never know what were actually offered in the option not chosen (that is, he or she will never know whether your message was true or not). Moreover, he or she will never know what to be paid to you according to the different options, nor will he or she know the actual amount you receive according to the selected option.

I choose to send (please circle one message):

Message 1 **Message 2**

Quiz:

1. In Message 1 " Option A will earn you more money than option B", who does "you" refer to?
   a. Own
   b. The other person
2. Is Message 1 true?
   a. Yes
   b. No
3. Is Message 2 true?

a. Yes

b. No

4. The other person will select between Option A and Option B after receiving the message you selected.

   a. True

   b. False

5. Will the other person know whether or not the message you send him is true?

   a. Yes

   b. No

6. How much will you receive, if the other person selects option A?

   a. $7

   b. $10

7. How much will the other person receive, if they select option A?

   a. $7

   b. $10

8. Will the other person know the amount to be paid to you according to option B, if they select option A?

   a. Yes

   b. No

Choice Condition

This part of the experiment describes a situation involving you and one other person.

In each decision you will be asked to select between two options, where the option that you choose will determine the payoffs for you and the other person.

There are 10 questions altogether. Consider each question separately. That is, treat each decision as if it alone will determine the final payoffs. Although the questions are hypothetical, please do your best to respond as if the decisions are for real money.

Example: In this example, we will ask you to choose one of the two possible payment options. The two options are:

Your choice will be executed with 80 percent of chance, while in the other 20 percent the alternative option will be implemented.

The other person will never know what sums were actually offered in the option not chosen. Moreover, he or she will never know the sums to be paid to you according to the different options, nor will he or she know the actual amount you receive according to the selected option.

I choose (please circle one option):

Option A **Option B**

Quiz:

1. Which option will give you more money on average?
   a. Option A
   b. Option B
2. Which option will give the other person more money on average?
   a. Option A
   b. Option B
3. Will the other person know how much you receive?
   a. Yes
   b. No

A supplementary methods checklist is available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Daniel Auerbach, Zac Robertson, and Clay Clayworth for assistance with data collection, analyses, and lesion reconstruction. This research was supported by the National Institutes of Health (R01 MH098023 to MH; R01 MH087692 to PC; R01 DA036017 to BKC; and R01 NS21135 to RTK), Hellman Family Faculty Fund (MH), VA ORD RR&D (D7030R to BKC) and the Nielsen Corporation (RTK).

References (Main Text)

1. Sally D. Rationality and society. 1995; 7:58–92.
2. Gneezy U. The American Economic Review. 2005; 95:384–394.
3. Greene JD, Paxton JM. PNAS. 2009; 106:12506–12511. [PubMed: 19622733]
4. Nunez JM, Casey B, Egner T, Hare T, Hirsch J. Neuroimage. 2005; 25:267–277. [PubMed: 15734361]
5. Christ SE, Van Essen DC, Watson JM, Brubaker LE, McDermott KB. Cerebral cortex (New York, NY : 1991). 2009; 19:1557–1566.
6. Spence SA, et al. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2004; 359:1755–1762. [PubMed: 15590616]
7. Somerville LH, Casey B. Current opinion in neurobiology. 2010; 20:236–241. [PubMed: 20167473]
8. Sodian B, Frith U. Journal of child psychology and psychiatry. 1992; 33:591–605. [PubMed: 1577901]
9. Searcy WA, Nowicki S. The Evolution of Animal Communication: Reliability and Deception in Signaling Systems. 2010
10. Camerer C. Behavioral game theory: Experiments in strategic interaction. 2003
11. Figner B, et al. Nature neuroscience. 2010; 13
12. Hare T, Camerer CF, Rangel A. Science. 2009; 324:646–648. [PubMed: 19407204]
13. Sip KE, Roepstorff A, McGregor W, Frith CD. Trends in cognitive sciences. 2008; 12:48–53. [PubMed: 18178516]
14. Rosano C, et al. Biological psychiatry. 2005; 57:761–767. [PubMed: 15820233]
15. Tapert SF, et al. Psychopharmacology. 2007; 194:173–183. [PubMed: 17558500]
16. Koenigs M, et al. Nature. 2007; 446:908–911. [PubMed: 17377536]
17. Krajbich I, Adolphs R, Tranel D, Denburg N, Camerer CF. J Neurosci. 2009; 29:2188–2192. [PubMed: 19228971]
18. He BJ, et al. Neuron. 2007; 53:905–918. [PubMed: 17359924]
19. Greene J, Sommerville R, Nystrom LE, Darley JM, Cohen J. Science. 2001; 293:2105–2108. [PubMed: 11557895]
20. Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E. Science. 2006; 314:829–832. [PubMed: 17023614]

References (Online Methods)

21. Rorden C, Brett M. Behavioural neurology. 2000; 12:191–200. [PubMed: 11568431]
22. Crawford V. Journal of Economic theory. 1998; 78:286–298.
23. Charness G, Rabin M. Quarterly Journal of Economics. 2002; 117:817–869.
24. Andreoni J, Miller J. Econometrica. 2002; 70:737–753.

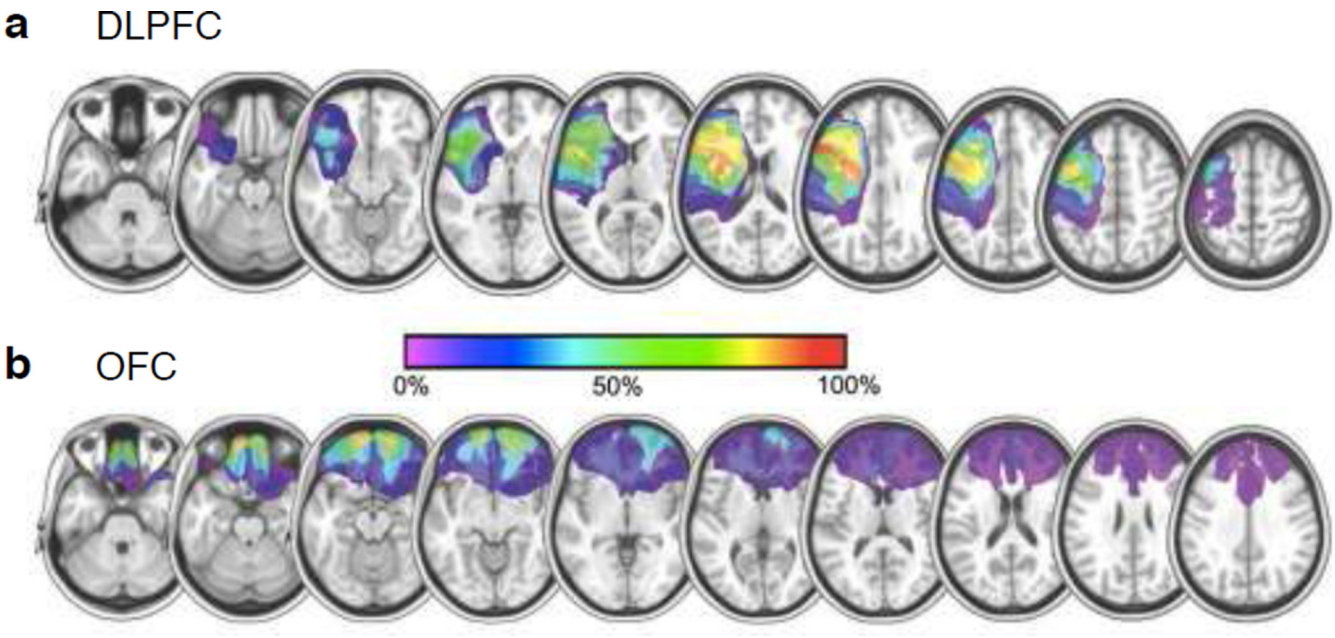


Figure 1.

Lesion reconstruction. Structural MRI slices illustrating the lesion overlap across the two patient groups. **(a)** For the DLPFC group (n = 6), mean lesion volume was 125.76 cm^3 and maximal cortical lesion overlap (>50%) was in the Brodmann areas 6, 8, 9 and 46, encompassing portions of the middle and superior frontal gyri in all patients. All dorsolateral prefrontal cortex lesions (5L; 1R) were shown overlapped to the left hemisphere for comparison purposes. For lateralized and individual reconstruction see Supplementary Fig. 1-2 and Supplementary Table 1. **(b)** For the orbitofrontal cortex group (n = 7), mean lesion volume was 72.29 cm^3 and maximal cortical lesion overlap (>50%) was in Brodmann areas 10, 11, and 47, centered in the OFC and including portions of inferior and superior frontal gyri in some patients. See Online Methods for details.

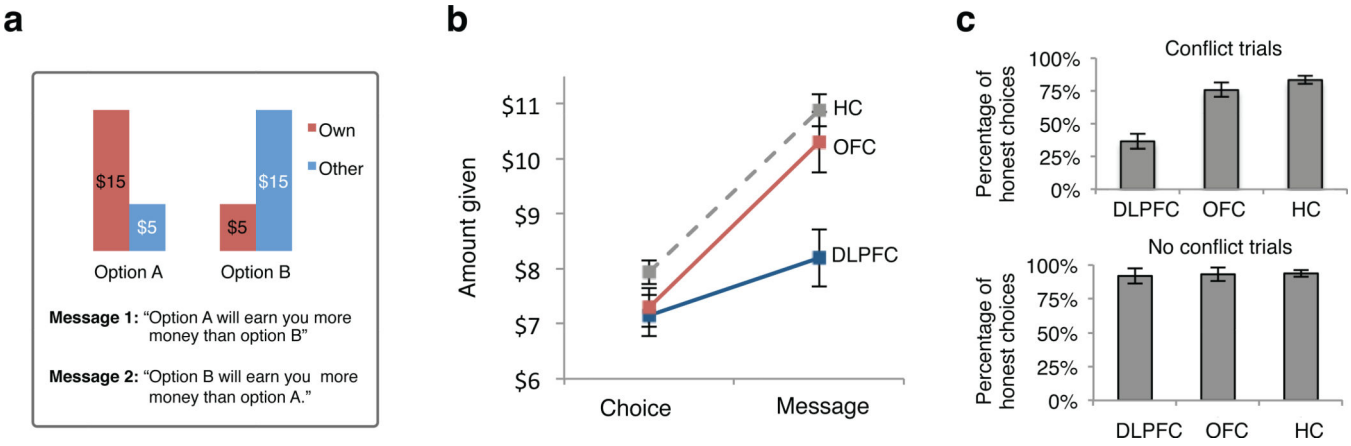


Figure 2.

Experimental paradigm and behavioral results. **(a)** Experimental paradigm. In the Message condition, the participant in the role of the signaler is presented with two options, A and B, associated with different monetary consequences. For example, Option A corresponds to $15 to the participant and $5 to an anonymous signal recipient, i.e., ($15, $5), and Option B corresponds to ($5, $15). There are furthermore two actions available to the participant in the form of two statements describing the monetary consequences of the options to the recipient. Specifically, the participants must choose between sending a truthful message (Message 2) that sacrifices economic self-interest in favor of honesty, or a false message (Message 1) that satisfies self-interest at the expense of being honest. See Online Methods for details. **(b)** Amount given. In the Choice condition, all cohorts gave similar amounts to the recipient (Healthy Comparison: $7.44 ± .22; DLPFC: $6.65 ± .38; OFC: $6.79 ± .35; Kruskal-Wallis test, p > .10, two-tailed). In the Message condition with identical monetary consequences but with the addition of honesty concerns, healthy participants increased giving by $2.94 ± .44. In contrast, DLPFC cohort's giving increased by less than half this amount ($1.05 ± .43), and significantly lower than those of the healthy comparison cohort (Wilcoxon rank sum test, $p < .001$, two-tailed). Finally, OFC participants were nearly identical to healthy participants ($3.01 ± .55; Wilcoxon rank sum test, $p > .50$, two-tailed), and significantly different from DLPFC participants (Wilcoxon rank sum test, $p < .001$, two-tailed). **(c)** Conflict and no conflict trials. On trials in the Message condition where honesty motives conflicted with those of self-interest (Top), DLPFC patients made a significantly lower proportion of honest choices (36.7% + 5.75%) compared to OFC and healthy comparison cohorts (OFCs: 75.7% + 5.44%; healthy participants: 83.3% + 3.00%; Fisher's exact test, $p < .01$ for both, two-tailed). In contrast, on trials where conflict was absent (Bottom), no significant differences existed between cohorts (Fisher's exact test, $p > .20$, two-tailed). All error bars indicate SEMs.

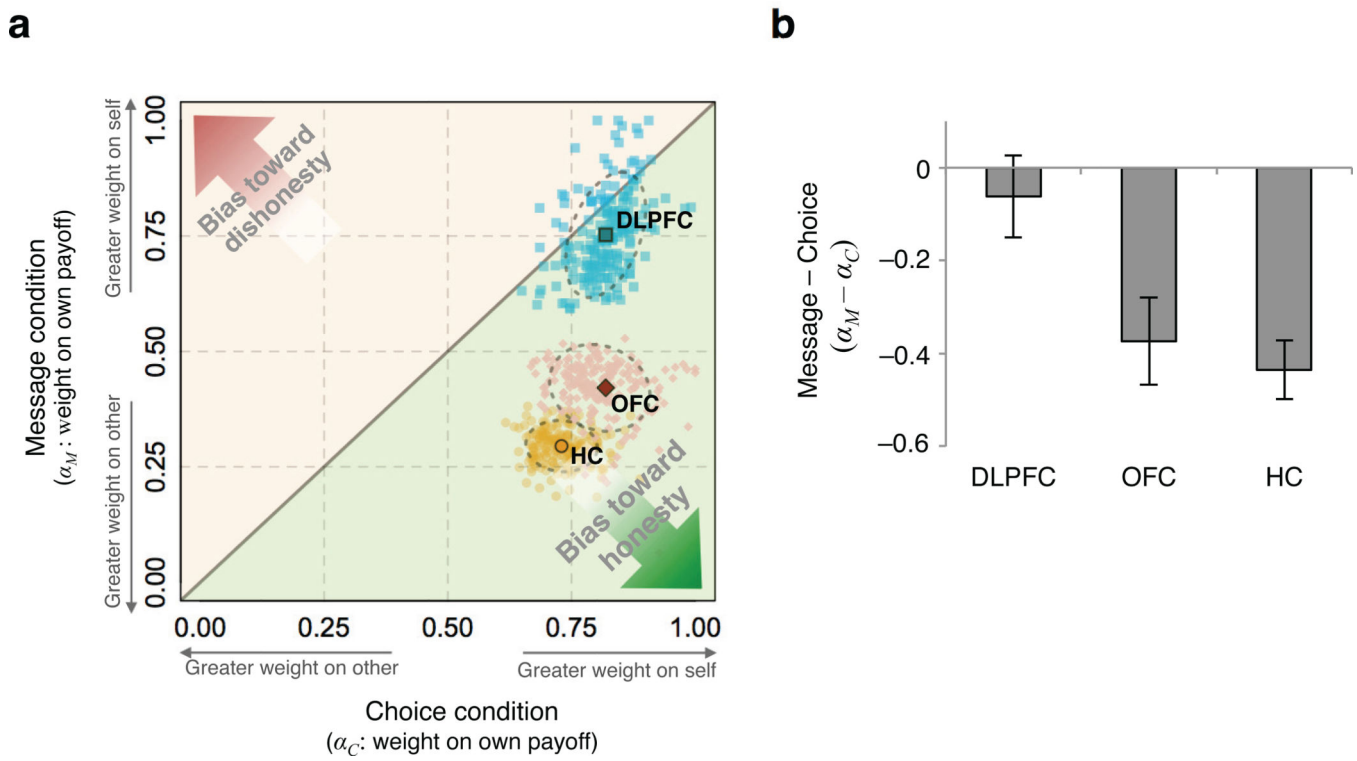


Figure 3.

Computational modeling. **(a)** Green shaded region captures willingness to sacrifice own payoffs to send the true message, i.e., bias toward honesty, where weight on self-interest in the Message condition (α_M) is reduced relative to the Choice condition (α_C). Conversely, red shaded region captures willingness to sacrifice own payoffs to send the false message, i.e., bias toward dishonesty, where α_M is greater than α_C. All cohorts placed similar weights on one's own payoff in the Choice condition (DLPFC: . 82 ± .05, OFC: . 79 ± .07 and healthy comparison: . 73 ± .05). In the Message condition, OFC and healthy participants showed a significant reduction in weight on own payoff, whereas DLPFC participants did not differ significantly between the two conditions (DLPFC: . 75 ± .09; OFC: .43 ± .06; and healthy comparison:. 29 ± .04). Solid points represent parameter estimates and smaller points represent bootstrap pseudo-sample estimates. Dashed ellipses correspond to bootstrapped standard errors. **(b)** Taking paired-wise differences in pseudo-sample estimates of α_M and α_C, OFC and healthy participants showed significantly lower weight on own payoff in the Message condition as compared to the Choice condition ($p < .01$, two-tailed), whereas the DLPFC cohort did not exhibit a significant difference ($p > .05$, two-tailed; all error bars indicate bootstrap standard errors).